Title:
Google Search Interests and New Cases of COVID-19 in Bangladesh: A Vector Autoregression Analysis for Disease Surveillance

Monir Uddin Ahmed[1], Mazbahul G. Ahamad[2] and Md Mahedi Hasan[3]

(**corresponding author**), Department of Economics, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh. E-mail: monir-eco@sust.edu

[2] University of Nebraska–Lincoln, 140 Keim Hall, 1825 N 38th St, Lincoln, NE 68583, USA.
E-mail: mahamad2@unl.edu

[3] Department of Mathematics and Statistics, Washington State University, WA 99164, USA.
E-mail: mdmahedi.hasan@wsu.edu

Running Title: COVID and VAR

Abstract

Background: The use of Google search engine has been widely used in the public health-related concerns. Previous studies found that Google search trends (GST) can predict influenza, mortality, Zika epidemics, Ebola etc. This study examines the relationship between timing of Coronavirus-related Google search trends, lockdown, and new cases of COVID-19 in Bangladesh.

Methods: We use national-level Google search trend data to examine whether the timing of Google search terms, i.e., their lag effects are associated with actual COVID-19 new cases from March 2, 2020, to December 7, 2020. We examine the effects of search terms (facemasks, handwash, n95) on the actual COVID-19 new cases using the vector autoregression (VAR) model.

Results: Our general recursive vector autoregression model shows that search on facemasks and hand-wash can potentially decrease the risk communication of COVID-19 new cases. We find that the search on facemasks can substantially reduce that risk in the sense that search can increase the use of facemasks. We also examine the lag effect of lockdown and find that the effects are not sizeable on the risk communication because their lag-effects are different. The results of the impulse-response functions show that among the protective measures, lag effects of facemasks can substantially decrease the future risk communication of COVID-19 new cases.

Conclusion: Because wearing facemasks can substantially reduce the risk of COVID-19 new cases, the government can utilize the Google search trends related to COVID-19 to disseminate the preventive information on COVID-19 and thus minimize the new cases and deaths.

Keywords: Google search trend; Relative search volume; Risk communication; vector autoregression; Bangladesh.

Introduction

Over the past few months, the Google search engine has been widely used for public health-related concerns about local transmission of COVID-19 in almost 200 countries (Lai et al., 2019). Recent studies have found that Google search trends (GST) can predict influenza, mortality, Zika epidemics, Ebola, dementia, and dengue, among others (Yeh and Yeh, 2019; Kang et al., 2013; Milinovich et al., 2015; Teng et al., 2017; Wang et al., 2015; Rangarajan et al., 2019, Husnayain et al., 2020; Schnoell et al., 2020). Given the recent upsurge of COVID-19, the application of GST is critical for identifying epidemiological trends, exploring risk communication, and predicting human behavior regarding epidemics and pandemics, responses to different kinds of health measures, and so on.

In Bangladesh, almost 5 million confirmed cases and more than 7,000 deaths have occurred, with a mortality rate of 1.46% (Directorate General of Health Services, 2020). The high infection rate ("Bangladesh sees no drop", 2020), symptoms, risk communication, and treatment of COVID-19 require widespread information because knowledge of prevention of the disease is scant. Studies have found that GST are very much related to conventional epidemiological surveillance (Cervellin, 2017). This easy and cost-effective way of monitoring human behavior regarding COVID-19 can be used for public health policies. In this study, we analyze whether the GST for facemasks and handwash can affect new COVID-19 cases in Bangladesh. We analyze the timing of searches for these terms that may affect COVID-19 risk communication. The findings of this study can be used for public health surveillance for COVID-19-related risk communication and deaths.

Methods

In this study, a vector autoregression (VAR) model was used to explore the relative search volume of Google trends (Google, 2020) on facemasks, handwash, and lockdown on new COVID-19 cases. We filtered the search to within Bangladesh from March 02, 2020, to December 7, 2020. We examined whether these search variables could predict COVID-19 infection. Our VAR(p) model included four weekly variables in the following order: new COVID-19 cases (COVID), facemasks (FACE), handwash (HAND), and lockdown (LOCK). To determine the keywords, we applied the glossary of daily reports from the Directorate General of Health Services Bulletin and daily newspapers. We then performed the principal component analysis (Wold et al., 1987) for similar terms and selected the aforementioned variables.[1] We also included a dummy variable for lockdown to see the direct intervention in COVID-19 infection. It also acted as a control variable. This defined a time-dummy where we imposed the value 1 if government-imposed lockdown existed, otherwise zero. Notably, after May 30, 2020, the government withdrew the lockdown to ease economic activities. After estimation of VAR, we performed the Granger causality test and impulse-response function. The Granger causality test shows whether all the lagged items of a variable influence the present value of the other variable(s). An impulse-response function describes the reaction of the variable of interest along a specified time horizon after a shock in a given moment.

Results and Discussion

We first estimated the level VAR using $p = 2$ based on the lag-length criterion. Table 1 shows that COVID depended on its lag period. This means that if we do not impose any restriction on the other measures, prior numbers of cases may produce more new cases in the upcoming period. Kucharski et al. (2020) estimated that the reproduction number will rise if the protective measures are not employed. Ceylan (2020), Roy et al. (2020), Alzahrani et al. (2020), and many other studies have found that new cases in the previous period can predict future cases. We also found that FACE and HAND can decrease the number of new COVID-19 cases. It is assumed that a high search volume is related to the inspiration

1 Principal component analysis results are available on request.

to use and purchase facemasks and handwash materials. These purchases and use have a significant effect on risk communication. This finding resembles that of Lin et al. (2020). We found that for both FACE and HAND, wearing facemasks and washing hands in the previous week may reduce the number of new cases. We also found that LOCK, the government-imposed lockdown, had an inconclusive effect on new cases because its lag effects are different. Diagnostic checks and autoregressive roots suggest that the level VAR estimation result is stable.

The Granger causality test results show that the previous values of one variable help to predict the other variable of its current state and beyond.[2] Table S3 shows that FACE, HAND, and LOCK Granger-cause COVID, and the results are significant. Our results suggest that past values of these variables can predict COVID significantly. We did not find bidirectional causality between COVID and FACE, or COVID and HAND.

In the case of the impulse-response function, we examined whether the shocks in GST (innovations in search trends in the case of COVID-19 risk communication) can affect new COVID-19 cases. We intended to check how this shock persisted over the periods. Ahamad et al. (2020) showed that the trends of GST have been declining. We identified the responses of COVID to shocks of GST variables. Panels A–C of Figure 1 show the responses of COVID to the shocks to FACE, HAND, and LOCK, respectively. We found that shock in the search of FACE can generate an innovation to COVID, which means that a shock in the search of FACE decreases the new COVID-19 cases. Panel A shows that new cases may decrease if people wear facemasks for the last two weeks, which is reflected in GST. This suggests that GST for facemasks has a significant effect on the risk communication of COVID-19, and the search timing is noteworthy. Panel B shows the same results for HAND. It shows the ineffectiveness of HAND on the reduction of new cases because of low levels of accessibility, insufficient knowledge on use, inappropriate distribution, and marketing of hand sanitizers and handwash materials at the beginning of the risk. Panel C shows that LOCK also increases the risk of COVID-19 initially. We found that the imposition of lockdown did not substantially reduce the risk, which was a common feature in most developing countries.

Conclusion

In this study, we explored whether GST related to COVID-19 affects new COVID-19 cases. Our results showed that searches for facemasks and handwash are negatively associated with and can potentially decrease the numbers of new COVID-19 cases. The results of the impulse-response functions showed that among the protective measures, lag effects of facemasks can substantially decrease the future risk communication of new COVID-19 cases. We also found that lockdown had an inconclusive effect on new cases. This urges public health policymakers to use GST to understand COVID-19 infections and deaths. The decreasing trend of searches for facemasks and other measures indicate that people are neglecting to wear masks. This may increase new cases in the winter. Because wearing facemasks can substantially reduce the risk of new COVID-19 cases, the government can use GST to disseminate COVID-19 preventive information and thus minimize new cases and deaths.

Conflict of interest

No competing interest declared.

Funding source

None.

Ethical approval

Ethical approval is not required for deidentified open data.

2 See Table S3 also in supplementary.

References

Ahamad, M. G., Ahmed, M. U., & Uden, D. R. (2020). Public Search Interests related to COVID-19: Insights from Google Search Trends in Bangladesh. *medRxiv*.

Alzahrani, S. I., Aljamaan, I. A., & Al-Fakih, E. A. (2020). Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *Journal of infection and public health*, *13*(7), 914-919.

Bangladesh sees no drop in Covid infection rate (2020, August 08). The Daily Star. Retrieved from www.thedailystar.net.

Cervellin G, Comelli I, Lippi G. (2017). Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Glob Health*. 2017;7:185-189.

Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of The Total Environment*, 138817.

Directorate General of Health Services (2020). Coronavirus COVID-19 Dashboard, 2020. Directorate General of Health Services (DGHS), Web: http://dashboard.dghs.gov.bd/, (accessed December 27, 2020).

Google (2020). Google Trends 2020. https://trends.google.com (accessed December 27, 2020).

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424-438.

Husnayain, A., Fuad, A., & Su, E. C. Y. (2020). Applications of google search trends for risk communication in infectious disease management: A case study of COVID-19 outbreak in Taiwan. *International Journal of Infectious Diseases*.

Kang, M., Zhong, H., He, J., Rutherford, S., & Yang, F. (2013). Using google trends for influenza surveillance in South China. *PloS one*, *8*(1), e55205.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., ... & Davies, N. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases*.

Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. *Int J Antimicrob Agents*. 2020;55:105924.

Lin, Y. H., Liu, C. H., & Chiu, Y. C. (2020). Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. *Brain, Behavior, and Immunity*.

Lütkepohl, Helmut. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335-346.

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of applied econometrics*, *11*(6), 601-618.

Milinovich, G. J., Magalhães, R. J. S., & Hu, W. (2015). Role of big data in the early detection of Ebola and other emerging infectious diseases. *The Lancet Global Health*, *3*(1), e20-e21.

Rangarajan, P., Mody, S. K., & Marathe, M. (2019). Forecasting dengue and influenza incidences using a sparse representation of Google trends, electronic health records, and time series data. *PLoS computational biology*, *15*(11), e1007518.

Roy D., Tripathy S., Kar S.K., Sharma N., Verma S.K., Kaushal V. (2020). Study of knowledge, attitude, anxiety & perceived mental healthcare need in Indian population during COVID-19 pandemic. Asian J. Psychiatr. 2020

Schnoell, J., Besser, G., Jank, B. J., Bartosik, T. J., Parzefall, T., Riss, D., ... & Liu, D. T. (2020). The association between COVID-19 cases and deaths and web-based public inquiries. *Infectious Diseases*, 1-8.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1-48.

Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic perspectives*, *15*(4), 101-115.

Teng, Y., Bi, D., Xie, G., Jin, Y., Huang, Y., Lin, B., ... & Tong, Y. (2017). Dynamic forecasting of Zika epidemics using Google trends. *PloS one*, *12*(1), e0165085.

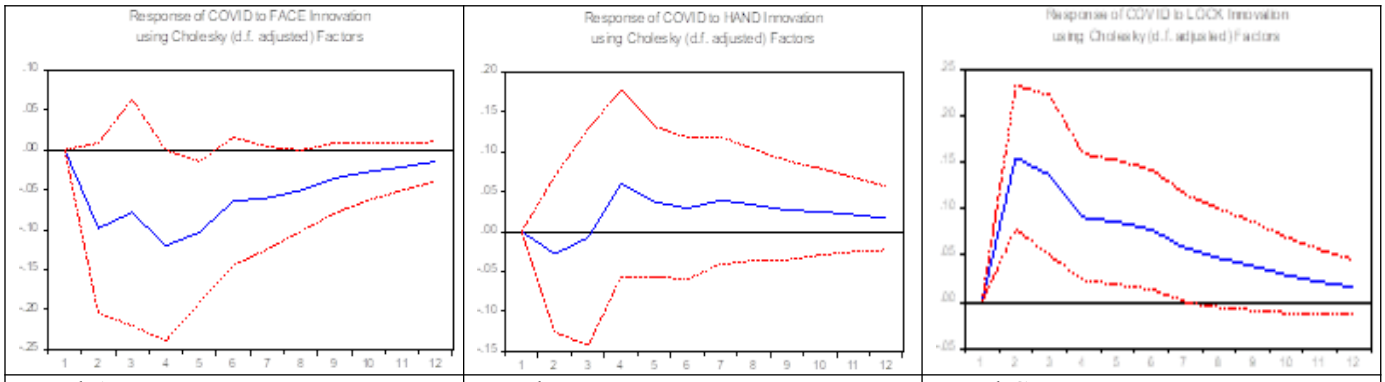
Wang, H. W., Chen, D. R., Yu, H. W., & Chen, Y. M. (2015). Forecasting the incidence of dementia and dementia-related outpatient visits with Google trends: evidence from Taiwan. *Journal of medical Internet research*, *17*(11), e264.

WHO (2020). WHO Coronavirus Disease 79 (COVID-19) Dashboard 2020. https://covid19.who.int/

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1-3), 37-52.

Yeh, F. C., & Yeh, C. H. (2019). Developing mortality surveillance systems using Google trend: A pilot study. *Physica A: Statistical Mechanics and its Applications*, *527*, 121125.

Figure and Tables



Panel A | Panel B | Panel C

Figure 1: Impulse-response function (Impulse: FACE, HAND, COVID; Response: COVID)

Table 1: Result of the estimation of VAR

Variable	**Lag**		
	0	**1**	**2**
Constant	1.631*** (2.80)		
COVID		0.944*** (6.57)	−0.091 (−0.83)
FACE		−0.006** (−2.29)	0.002 (1.05)
HAND		−0.007*** (−2.72)	0.002 (0.70)
LOCK		1.036***	−0.606***

		(5.21)	(−2.82)
Diagnostic Checks			
R-sq	0.98	Adj. R-sq	0.97
SBIC	17.569	Chi-sq	1662.041***
LM Test for Autocorrelation [Chi-sq(2)]	25.775 [0.06]	Normality of Residuals [Chi-sq(2)]	6.427 [0.04]

Dependent variable is COVID and the regressors are the lags of COVID, FACE, HAND and LOCK
Figures in parenthesis are z-statistics.
Figures in square brackets are p-values.
*, ** and *** represent 10%, 5% and 1% level of significance.

Table S3: Granger-causality test

Lag-length	Dependent Variable(Y)	Independent Variable(X)	Chi-sq Statistics	Decision
2	COVID	FACE	6.331** (0.04)	$HAND \rightarrow COVID$
2	COVID	HAND	8.662** (0.01)	$FACE \rightarrow COVID$
2	COVID	LOCK	34.72*** (0.00)	$LOCK \rightarrow COVID$

Chi-sq measures the Wald test statistics.
Figures in parenthesis represent p-values.
Lag-length selection is based on SBIC.
Null hypothesis of the test is that X does not Granger-cause Y.
** and *** represent 5% and 1% level of significance respectively.